Setwise methods for the invariant measure problem: convergence?

Rua Murray (University of Canterbury, NZ)

SON2012 (UNSW), September 6, 2012

Part I: Ulam’s method, convergence and quasicompactness

Part II: Variational approaches

Part III: A few examples

Part I

Basic setup

- $T : (X, m) \to (X, m)$, $m \circ T^{-1} \ll m$
- **Assume:** There exists $\mu \circ T^{-1} = \mu$, $\mu \ll m$
  so that
  - $\int_X g \circ T \, d\mu = \int_X g \, d\mu$ for all $g \in L^\infty(X)$
  - μ has a *density function* $\frac{d\mu}{dm}$
- **Questions**
  - when can Ulam's method be expected to give good results?
  - are there reasonable alternatives?

Ulam's method for invariant measure problem

- Suggested by Ulam in 1960
- $\mathcal{A}_n = \{A_1, \ldots, A_n\}$ a partition of X into n subsets
- seek *Histogram measures* μ_n where:

$$\frac{d\mu_n}{dm} \in \text{span}\{\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_n}\}$$

  and

$$(\mu_n \circ T^{-1})(A_j) = \mu_n(A_j) \qquad \text{for} \qquad j = 1, \cdots, n$$

- then, for all measurable E,

$$\mu_n(E) = \sum_{i=1}^{n} \mu_n(A_i) \frac{m(E \cap A_i)}{m(A_i)}$$

  $[\mu_n(A_i)]_{i=1}^n$—invariant prob. vector for a stochastic matrix (*eg* `eigs()`)

- **Question:** Does $\mu_n \to \mu$ as partition refined? how fast?

Alternative to setwise view – discretisation of L^1

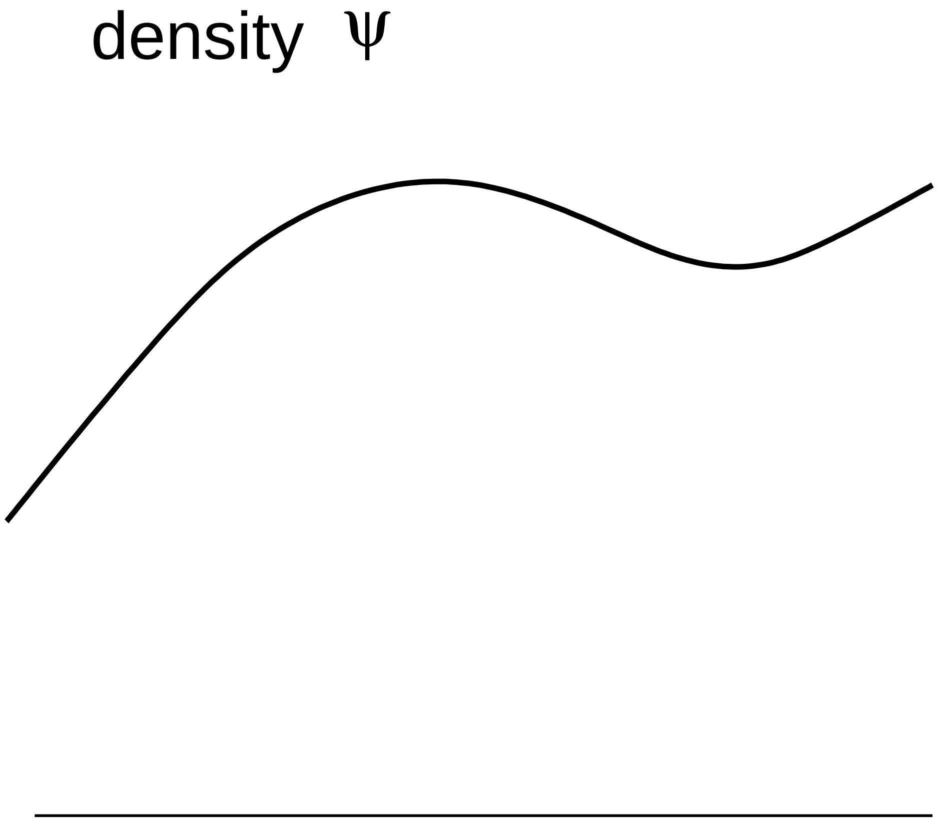


Figure 1: Begin with a density

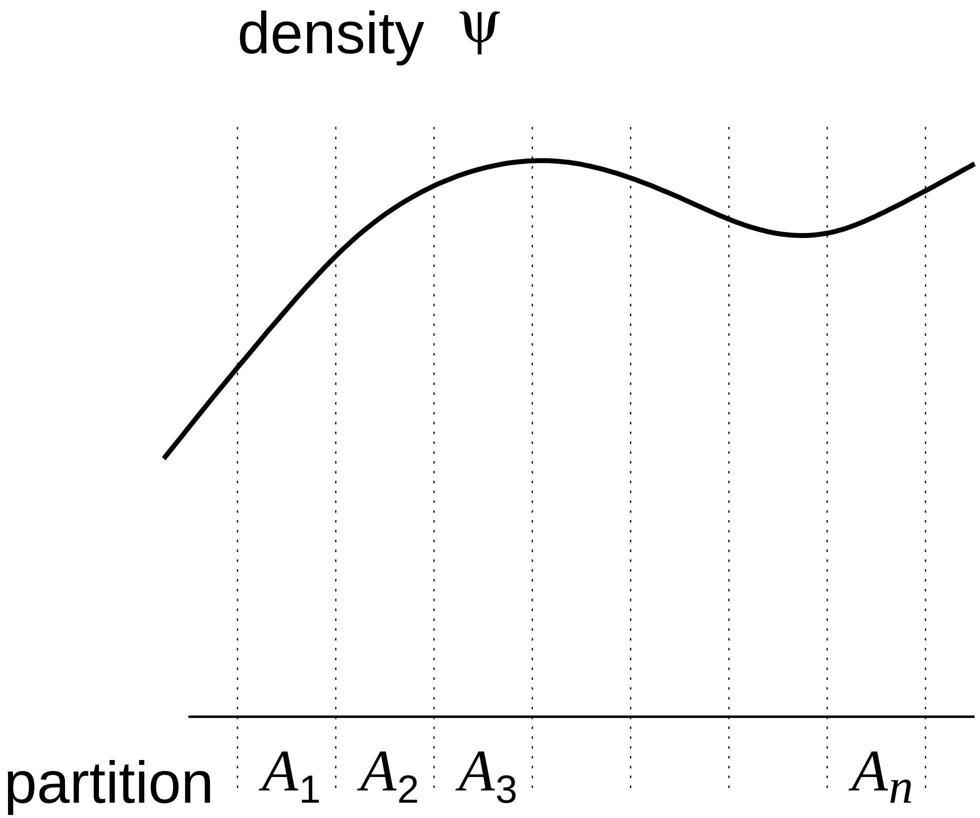


Figure 2: Choose a partition of X

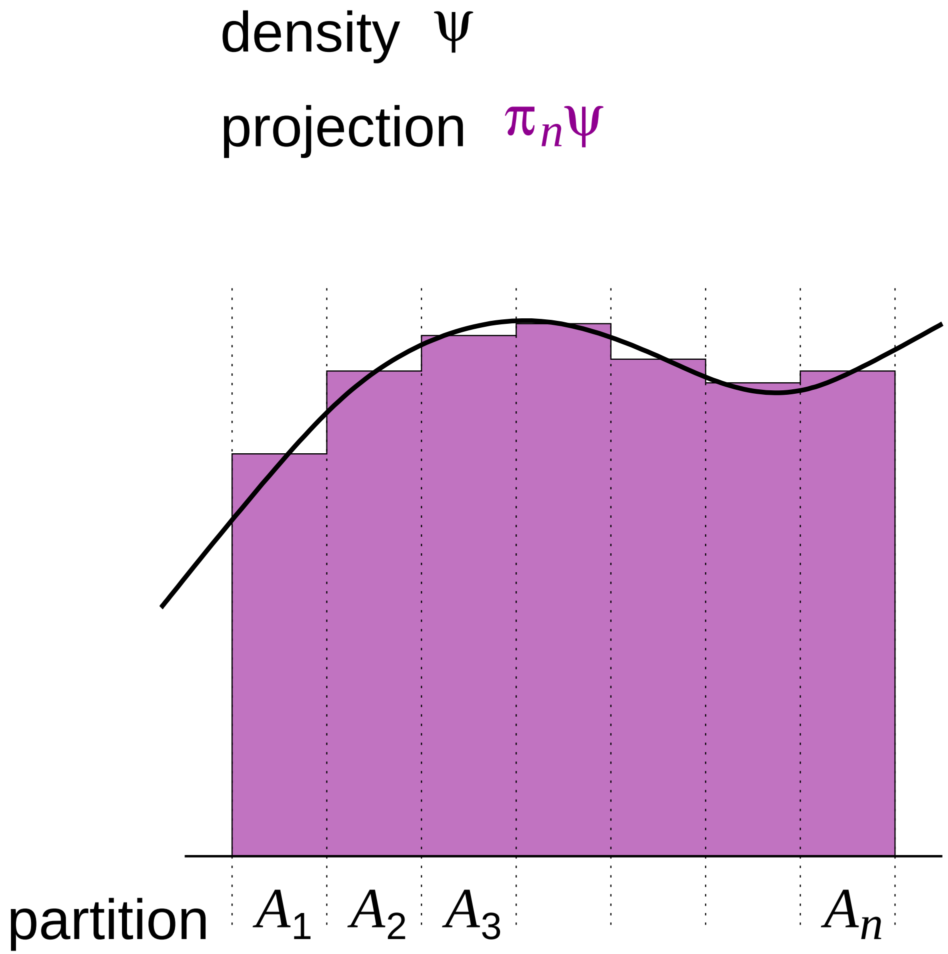


Figure 3: Project onto a *histogram*

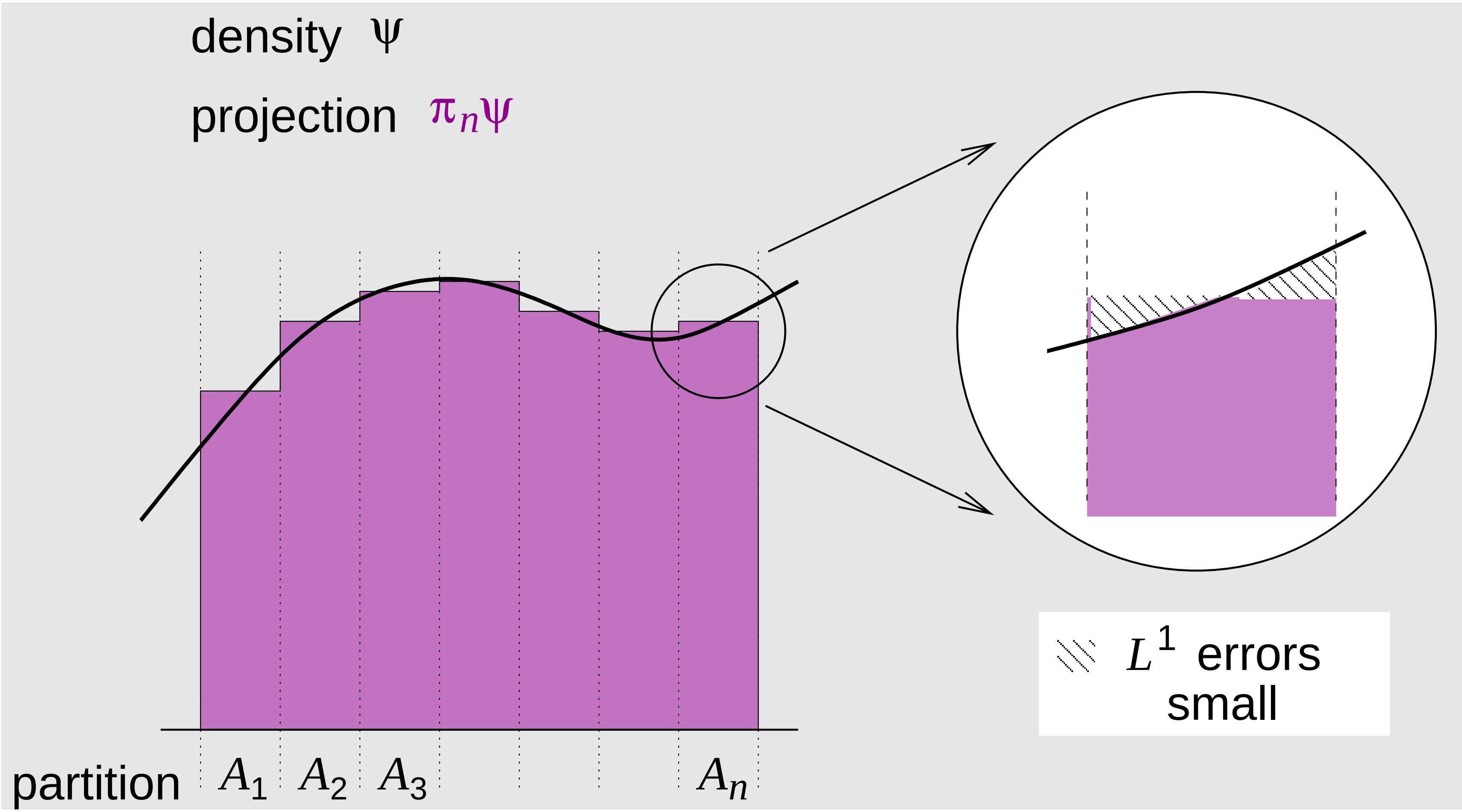


Figure 4: When ψ has good regularity, L^1–projection errors small

Observation: fixed points of Ulam's method are *approximately invariant*

$$\boxed{\mu_n(T^{-1}A_j) = \mu_n(A_j) \qquad j = 1, \cdots, n}$$

- But

  $$\mu_n(A_j) = \int \mathbf{1}_{A_j} \tfrac{d\mu_n}{dm}\, dm$$

  and

  $$\mu_n(T^{-1}A_j) = \int \mathbf{1}_{A_j} \circ T \tfrac{d\mu_n}{dm}\, dm$$

- so the densities $\frac{d\mu_n}{dm}$ from Ulam's method satisfy

  $$\int_X g \circ T\, d\mu_n = \int_X g\, d\mu_n$$

  for all $g \in$ span$\{\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_n}\}$

- Also
  - the Ulam measures have densities in span$\{\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_n}\}$
  - indeed, **any weak accumulation point of** $\{\mu_n\}$ is an invariant measure (many variants of this argument are published – [Lem3.1, Bose+M, *DCDS*06])
  - existence of *dynamically interesting* weak limits requires more work

Transfer operators

- The evolution of measures can be studied via their *density functions*
- Let ν be *absolutely continuous* wrt m (AC)
- Let $\psi = \frac{d\nu}{dm}$ and define $\mathcal{P} : L^1(X,m) \to L^1(X,m)$ by

$$\mathcal{P}\psi = \frac{d}{dm}(\nu \circ T^{-1})$$

  - called the *Frobenius–Perron* operator for T
  - $\mathcal{P}$ determines the evolution of density functions
  - $\|\mathcal{P}\|_{L^1} = 1$
  - fixed points of $\mathcal{P}$ are densities of ACIMs ($1 \in$ spec($\mathcal{P}$))
  - eigenfunctions of other $\lambda \in$ spec($\mathcal{P}$) related to *almost invariant sets* [Dellnitz, Junge, Froyland ·]

But: *are such eigenvectors accessed by Ulam's method?*

Lasota-Yorke and Li

Setting: Piecewise C^2 expanding interval maps

Lasota and Yorke (1973): $\exists \alpha \in (0,1), K < \infty$ s.t. for all $\psi \in BV$

$$\|\mathcal{P}\psi\|_{BV} \leq \alpha \|\psi\|_{BV} + K \|\psi\|_{L^1} \tag{LY}$$

$\Rightarrow \mathcal{P}$ has a fixed point in BV [compactness]

T-Y Li (1976): Ulam's method is a projection of fixed point problem for $\mathcal{P}$

- if $\psi \in L^1$ and $\mathcal{A}_n$ is the partition used for Ulam's method put

$$\pi_n(\psi) = \sum_{A_j \in \mathcal{A}_n} \left(\int_X \mathbf{1}_{A_j} \psi \, dm \right) \mathbf{1}_{A_j}$$

- the measures from Ulam's method satisfy

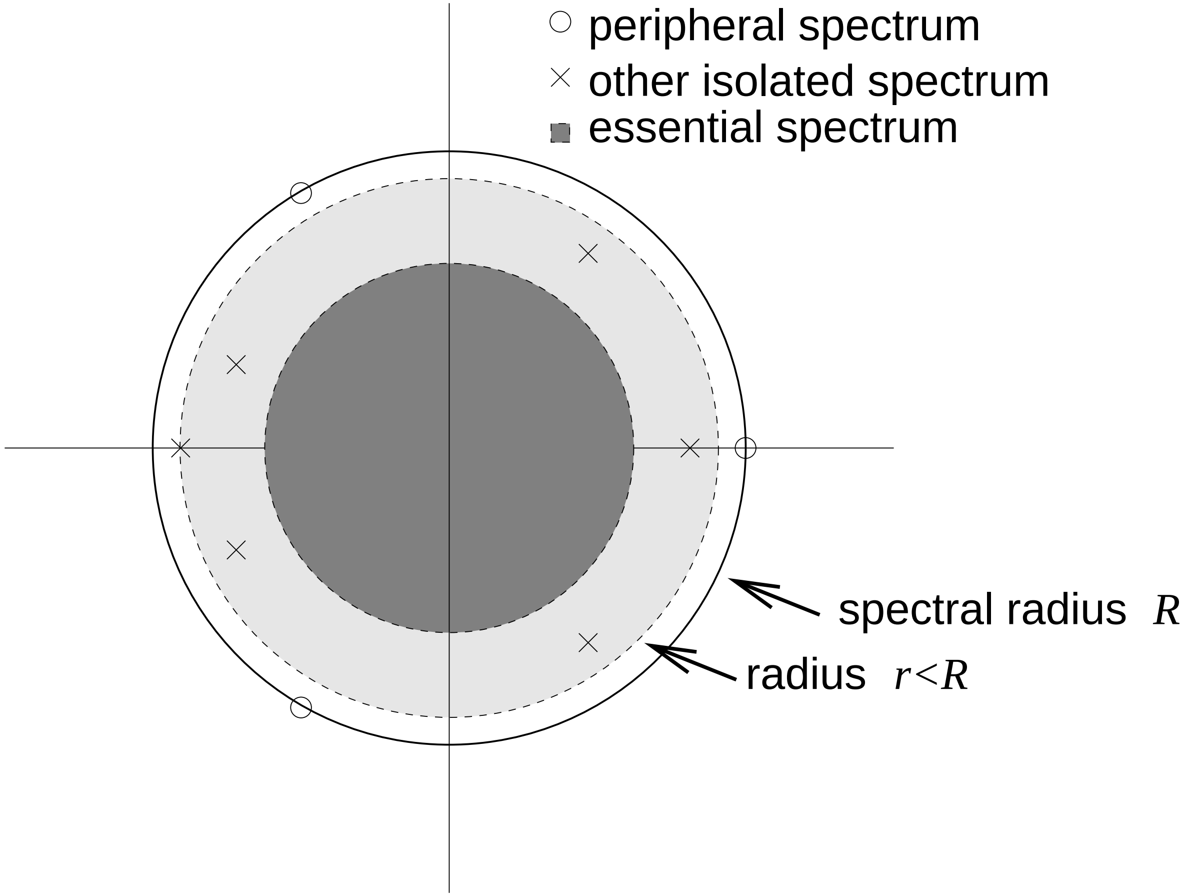
$$(\pi_n \circ \mathcal{P}) \frac{d\mu_n}{dm} = \frac{d\mu_n}{dm}$$

- since $\|\pi_n \psi\|_{BV} \leq \|\psi\|_{BV}$, (LY) holds with $\pi_n \circ \mathcal{P}$ replacing $\mathcal{P}$

$\Rightarrow$ **total-variation norm** convergence of μ_n to an AC invariant measure

Quasi-compactness

Inequalities like (LY) imply **quasi-compactness** of $\mathcal{P}$



Spectral perturbation

Keller (1982)

- (LY) + Ionescu-Tulcea and Marinescu ergodic theorem $\Rightarrow \mathcal{P}$ quasi-compact
- careful spectral perturbation argument gives *rate* in $O(\cdot)$ notation

Keller and Liverani (1999): arguments generalised, explicit constants given

- need “weak” $|\cdot|$ and “strong” $||\cdot||$ norms on $\mathcal{B}$
- $|\mathcal{P}^N| \le C_1\, M^N$, $||\mathcal{P}^N\psi|| \le C_2\alpha^N||\psi|| + C_3\, M^N\, |\psi|$ (generalising (LY))
- and these inequalities hold uniformly for a family $\{\mathcal{P}_\epsilon\}$ of operators

$\Rightarrow$ bounds on spectral projectors in terms of $\sup_{||\psi||=1} |(\mathcal{P}_\epsilon - \mathcal{P}_0)\psi|$

Summary: with very strong analytical control of T, one gets (LY) and strong bounds on errors in Ulam approximations

Part II [joint work with C Bose, UVic]

Problem: the above approach requires a lot of expansion and regularity on T (which is totally unrealistic in most applications)

We already know

1. Ulam measures are *approximately invariant*
   (expressed by a ‘dense’ collection of weak conditions)
2. Frobenius-Perron operators improve a fairly weak form of regularity $\cdots$

Markov operators increase relative entropy

- suppose $0 \le \psi_* = \mathcal{P}\psi_*$, $|\psi_*|_{L^1} = 1$ and put
$$H(f) = -\int_X f(x)\,\log(f(x)/\psi_*(x))\,dm(x)$$
where $\operatorname{supp}(f) \subseteq \operatorname{supp}(\psi_*)$
- by Jensen's inequality,
$$H(\mathcal{P}f) \ge H(f)$$
(and $H(f) \le H(\psi_*) = 0$ for $0 \le f$ with $|f|_{L^1} = 1$) see *eg* [Lasota-Mackey]
- since entropy maximisers are preferred by $\mathcal{P}$, and Ulam approximations are "as constant as possible", this may yield a "back-door" proof of convergence

...

Does Ulam select ‘entropy maximizers’ among approximately invariant measures?

- let $\mathcal{G}_n = \{g_1, \ldots, g_n\}$ be a collection of *test functions*
- $\mu \ll m$ is **approximately invariant up to** $\mathcal{G}_n$ if and only if

$$\int g \circ T \, \frac{d\mu}{dm} \, dm = \int g \, \frac{d\mu}{dm} \, dm \qquad \forall g \in \mathrm{span}(\mathcal{G}_n)$$

- let $\mathcal{F}_n$ be the densities of all such μ
- when $\mathcal{G}_n = \{\mathbf{1}_{A_i}\}_{A_i \in \mathcal{A}_n}$ the Ulam approximations have $\frac{d\mu_n}{dm} \in \mathcal{F}_n$
- if the partitions $\mathcal{A}_n$ are obtained by a sequence of subdivisions of X then $\mathcal{F}_{n+1} \subseteq \mathcal{F}_n$, span$(\cup_n \mathcal{G}_n)$ is weak-* dense in L^∞ and

$$\mu = \mu \circ T^{-1} \quad \Leftrightarrow \quad \frac{d\mu}{dm} \in \cap_n \mathcal{F}_n$$

- we (Bose + M) hoped that $\{\frac{d\mu_n}{dm}\}$ were the “most uniform” members of $\mathcal{F}_n$, and would converge to a finite-entropy invariant measure as $n \to \infty$

  $\cdots$ but this is usually false!

Selection of 'optimal' members of $\mathcal{F}_n$

Let

$$\Phi(f) = -\int_X f(x)\,\log f(x)\,dm(x)$$

Choose f_n to solve:

$$\max \Phi(f) \text{ subject to } f \in \mathcal{F}_n$$

Conveniently, $\Phi : L^1(X;\mathbb{R}) \to \mathbb{R}$ has the following properties:

- Φ is strictly concave
- Φ is weakly upper semi-continuous
- Φ has weakly compact upper level sets
- Φ is "Kadec" ($f_n \xrightarrow{\text{weak}} f$ + $\Phi(f_n) \to \Phi(f) \Rightarrow f_n \xrightarrow{L^1} f$)

the entropy functional provides a "fit for purpose" link, controlling regularity of approximately invariant measures, and assuring convergence

Convex duality and optimal approximate invariance

- membership $f \in \mathcal{F}_n$ can be rewritten in the following way:

$$\int_X [M^*\lambda]\, f\, dm = 0 \ \forall \lambda \in \mathbb{R}^n \qquad \text{and} \qquad \int_X f\, dm = 1$$

  where

$$M^*\lambda = \sum_{j=1}^{n} \lambda_j (g_j \circ T - g_j)$$

- using Fenchel duality, maximizing Φ subject to these constraints is equivalent to *unconstrained minimization* of

$$Q(\lambda, \lambda_0) := \int \exp(M^*\lambda + \lambda_0 \mathbf{1}_X - 1)\, dm - \lambda_0$$

  over $\mathbb{R}^{n+1}$

- an optimal λ exists if and only if

$$M^*\lambda \leq 0 \Rightarrow M^*\lambda = 0 \qquad \text{(*)}$$

- then, by solving $\frac{\partial Q}{\partial \lambda_i} = 0$,

$$f_n = e^{\lambda_0 - 1} \exp(M^*\lambda)$$

  is the entropy maximising element of $\mathcal{F}_n$

Summary so far

- choose partitions $\mathcal{A}_n$ which are successively finer
- $\mathcal{F}_n := \{0 \leq f \ : \ \int \mathbf{1}_{A_j} f \, dm = \int \mathbf{1}_{T^{-1}A_j} f \, dm, \forall A_j \in \mathcal{A}_n\}$
- weak limits of $f_n \in \mathcal{F}_n$ are automatically T-invariant
- if f_n solves $\max_{f \in \mathcal{F}_n, \|f\|=1} \Phi(f)$ then $\{\Phi(f_n)\}_{n \in \mathbb{N}}$ is a decreasing sequence
- Kadec property ensures that $\{f_n\}_{n \in \mathbb{N}}$ has an L^1-limit [Borwein-Lewis,'91]
- $\{f_n\}$ are calculated (via convex duality) as

$$f_n = e^{\lambda_0 - 1} \exp \left\{ \sum_{A_j \in \mathcal{A}_n} \lambda_j (\mathbf{1}_{A_j} \circ T - \mathbf{1}_{A_j}) \right\}$$

  [NOTE: functional form of f_n – not an element of span($\mathcal{G}_n$)]

- all of this can be accomplished provided condition ($*$) holds:

$$\sum_{A_j \in \mathcal{A}_n} \lambda_j (\mathbf{1}_{A_j} \circ T - \mathbf{1}_{A_j}) \leq 0 \Rightarrow \text{“}= 0\text{”}$$

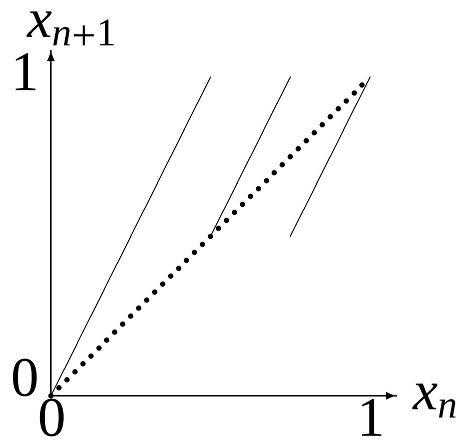
- (*) **can fail**, but we can *restrict to the largest domain X_n on which (*) holds*, obtaining convergence!

Dynamical interpretation of domain restriction

- When $h = \sum_{j=1}^{n} \lambda_j \mathbf{1}_{A_j}$, $M^*\lambda \leq 0$ implies $h \circ T \leq h$
- That is, the function h decreases along orbits
- So h is constant on “recurrent” parts of X
- Thus, places where $M^*\lambda < 0$ must be “transient”
- X_n is precisely the collection of “recurrent” components of the dynamics, as resolved by the partition $\mathcal{A}_n = \{A_1, \ldots, A_n\}$
- We can (quickly and easily) identify X_n almost everywhere (*w.r.t.* m), once the matrix $m(A_i \cap T^{-1}A_j)$ is known!

Part III

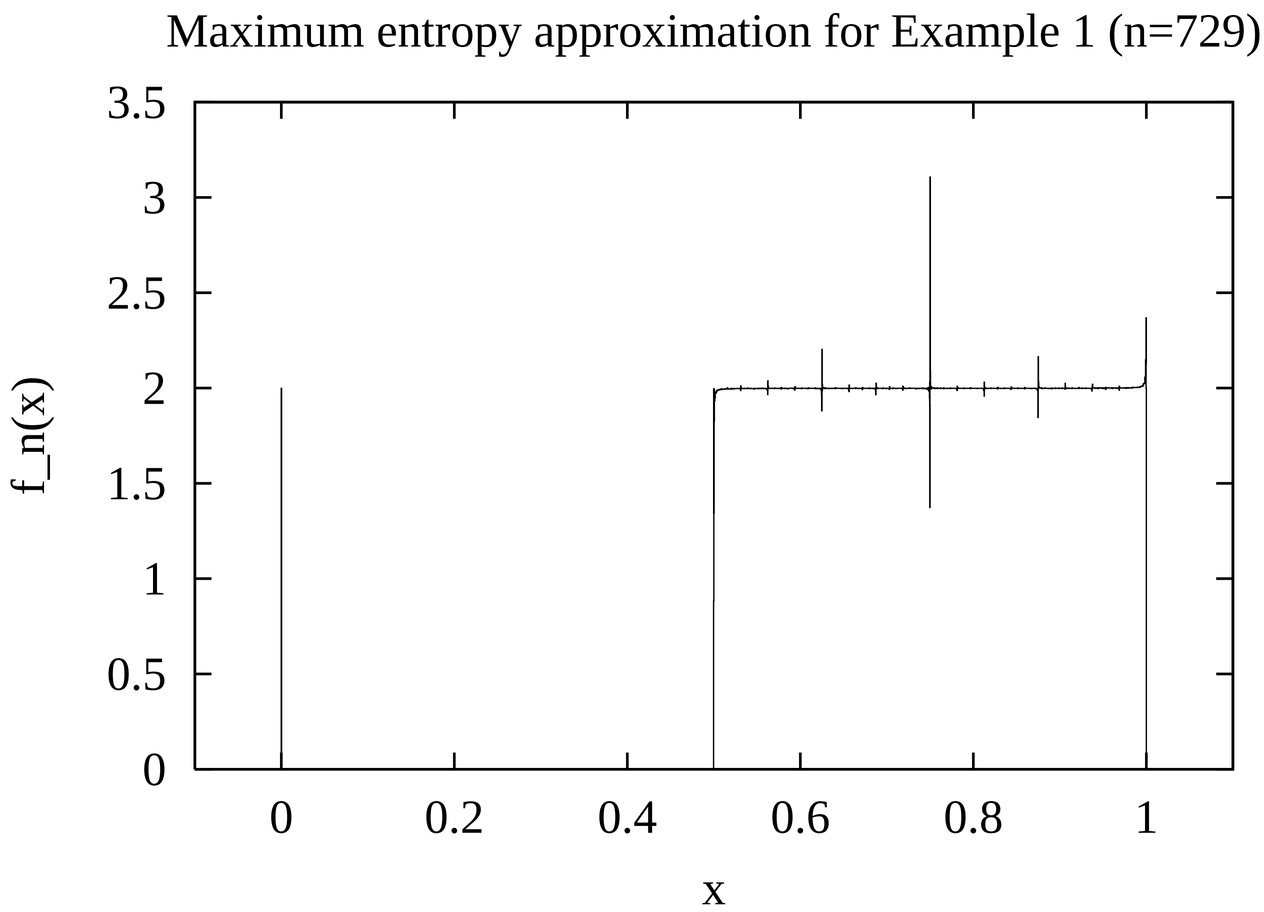
Example 1: a really simple map



- $T(x) = \begin{cases} 2\,x & \text{if } x \in [0, 1/2), \\ 2\,x - 1/2 & \text{if } x \in [1/2, 3/4), \\ 2\,x - 1 & \text{if } x \in [3/4, 1]. \end{cases}$
- The (unique) invariant density is $f_* = 2\,\mathbf{1}_{[1/2,1]}$
- With $n = 729$ uniform subintervals, we identify

$$X_n = \left[0, \frac{1}{2 \times 729}\right) \cup \left[\frac{364}{729}, 1\right]$$

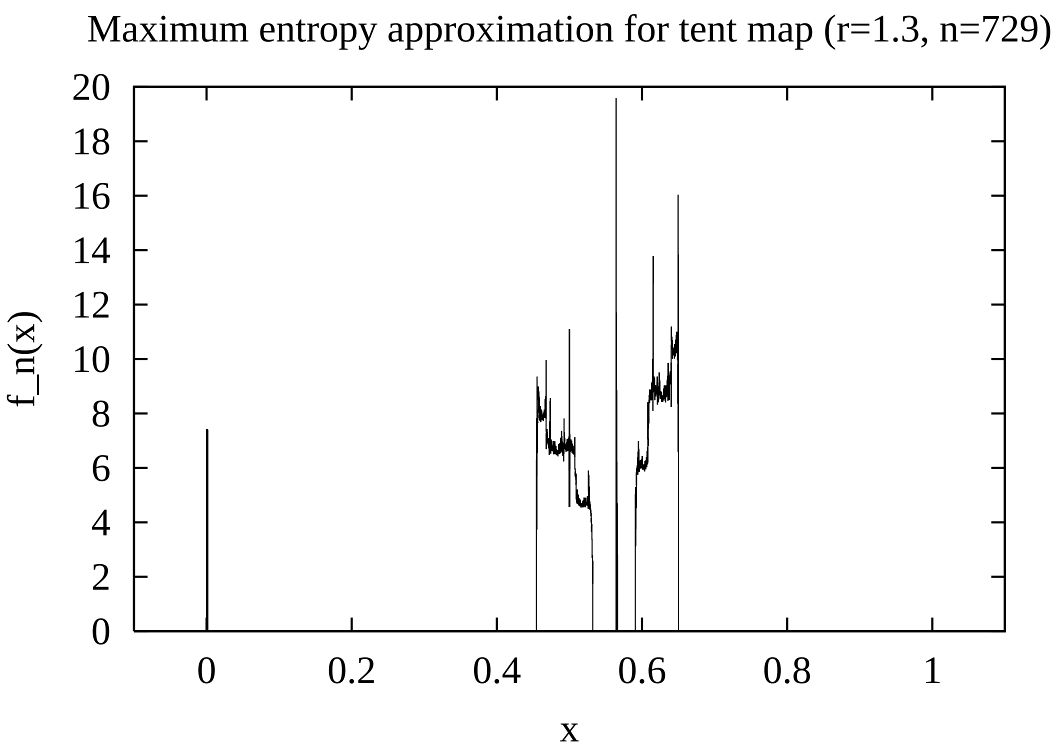
- The density approximation is depicted below



- X_n contains the support of f_* AND the repeller at 0
- Spikes are due to the discontinuity of f_* at $1/2$

Example 2: tent map with slope 1.3

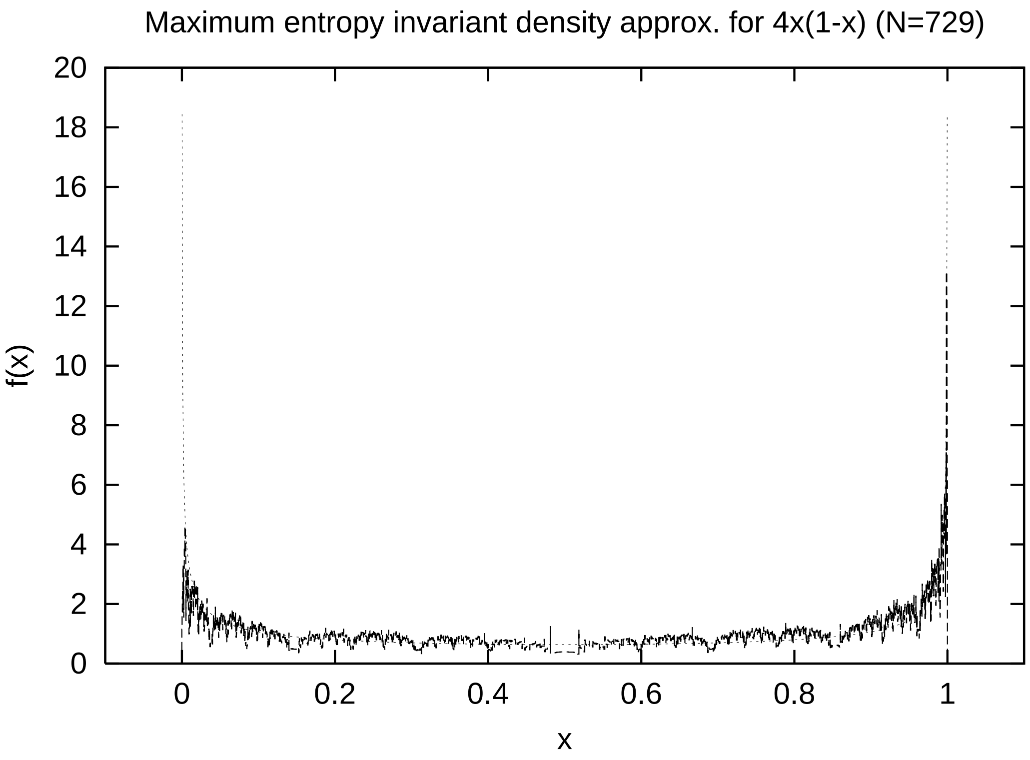
The density is supported on two intervals. Note the spikes corresponding to the unstable fixed points.



Example 3: full Logistic map

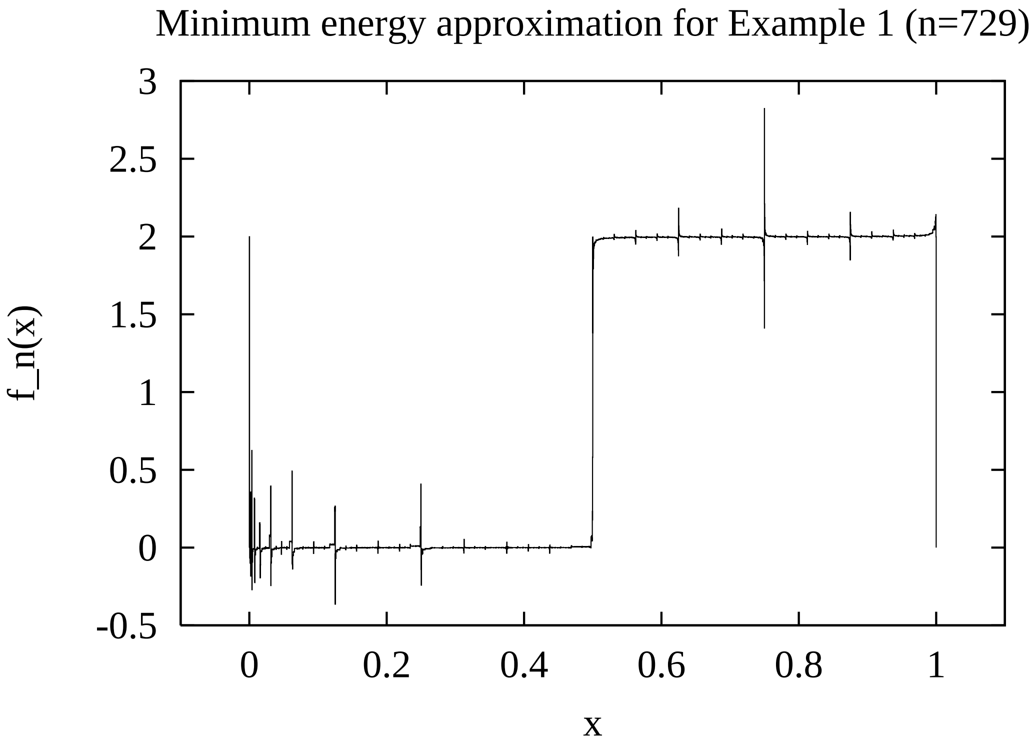
- $T(x) = 4\,x\,(1-x)$ on the interval $[0, 1]$
- The known invariant density is $f_*(x) = \frac{1}{\pi}\frac{1}{\sqrt{x\,(1-x)}}$

  Note that this density is fully supported, so $X_n = [0, 1]$



Example 4: Example 1 again, this time with 'minimum energy'

- solve $\min_{\mathcal{F}_n} \frac{1}{2}\|f\|^2_{L^2}$
- the same theory goes through for this convex functional
  - gain linearity of the optimality conditions
  - lose the need to do a domain restriction
  - can no longer be sure of positivity of the optimal solution!



Application to open systems

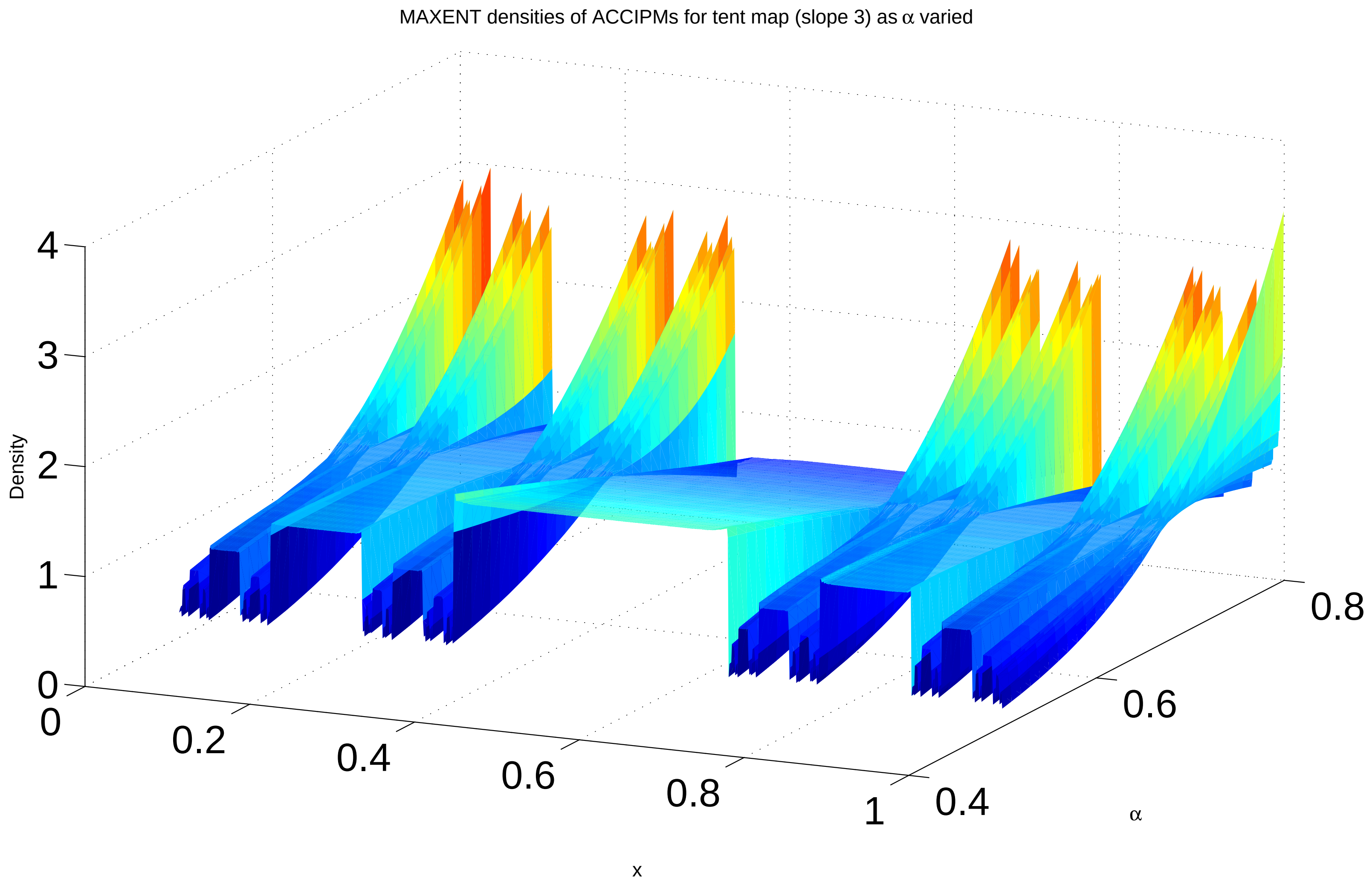
- suppose that T is 'open', so $T(X) \setminus X \neq \emptyset$
- instead of *invariant measures*, one seeks *conditionally invariant measures* solving
$$\mu \circ T^{-1} = \alpha \mu$$
for some $\alpha < 1$
- the **escape rate** of μ from the *repelling survivor set* is $-\log \alpha$
- one can build conditionally invariant measures for a range of α
- the same MAXENT analysis goes through
- conditionally invariant measures are supported on the repeller and its unstable manifold
- a domain restriction is sometimes required
  - essentially the same analytical condition
  - now related to "backwards transience"
    (those parts of X with no backwards orbits)

Example 5: tent-map with slope 3

- the map is familiar:

$$T(x) = \begin{cases} 3\,x & x < 0.5 \\ 3\,(1-x) & x > 0.5 \end{cases}$$

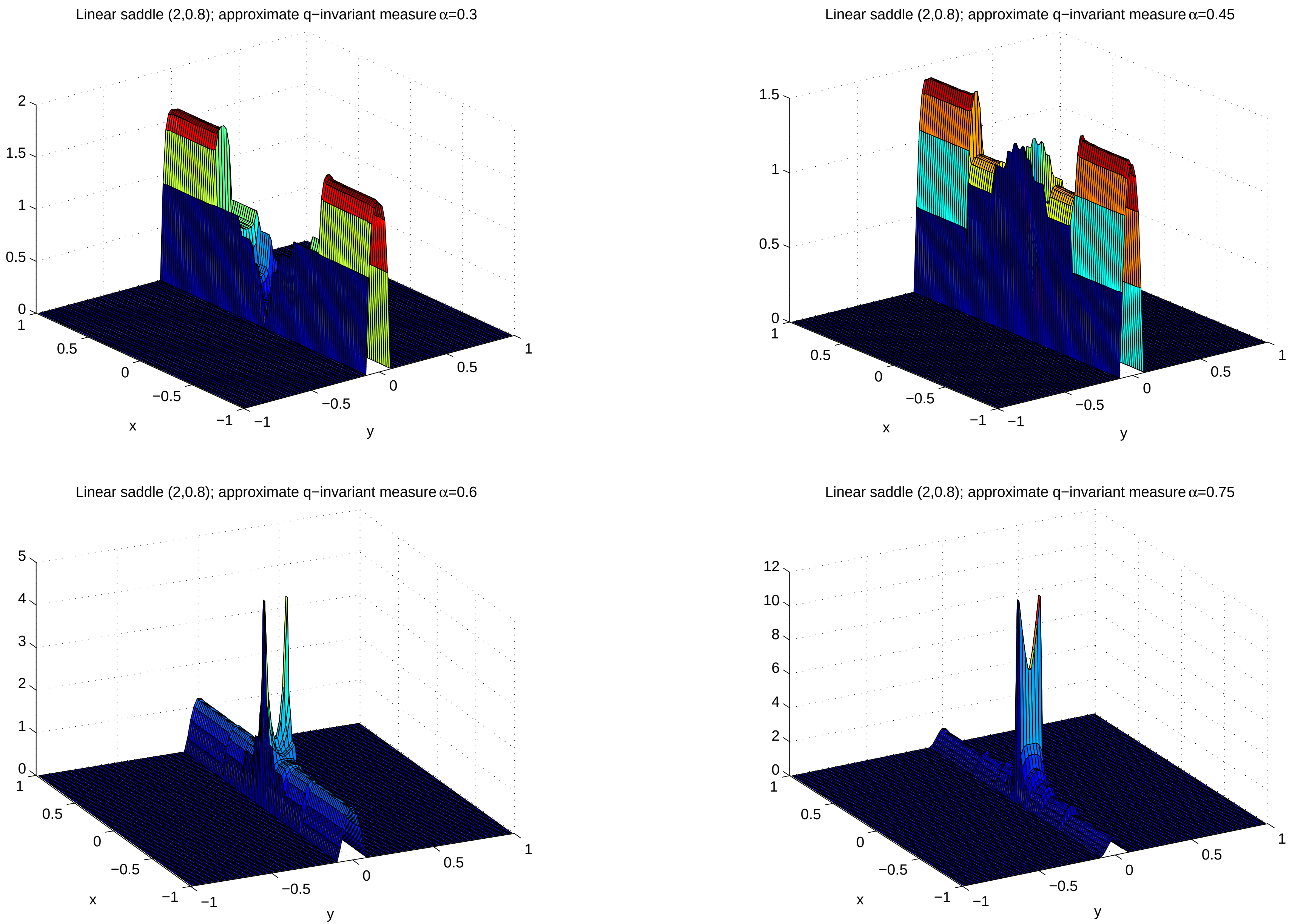
- the “hole” $(1/3, 2/3)$ escapes from $[0,1]$ under one iterate of T
- the “natural” conditionally invariant measure is Lebesgue; $f_* = 1$, with $\alpha = 2/3$
- the invariant repeller is the usual “middle thirds Cantor set”
- the MAXENT method can be tuned to produce a “most uniform” approximated ACCIPM for each $\alpha \in (0,1)$
- computations done with test functions $\{g_j = \mathbf{1}_{[(j-1)/1000, j/1000)}\}_{j=1}^{1000}$


MAXENT densities of ACCIPMs for tent map (slope 3) as α varied
Density
0
1
2
3
4
0
0.2
0.4
0.6
0.8
1
x
0.4
0.6
0.8
α

Example 6: trial escape computations for a linear saddle

Non-rigorous: *there is no absolutely continuous conditionally IM*

- $X = [-1, 1]^2$, m=Area
$$T(x, y) = (2\,x, 0.8\,y)$$
- The only invariant measure is δ-measure at 0
- all conditionally invariant measures lie on the unstable manifold to 0 (x-axis)
- a selection of these can be approximated by tuning $\alpha \in (0, 1)$
- $\{g_j\}$ are characteristic functions on subrectangles in a 100×100 grid
- computations are implemented with a few dozen lines of MATLAB code


Linear saddle (2,0.8); approximate q−invariant measure α=0.3
x
y
Linear saddle (2,0.8); approximate q−invariant measure α=0.45
x
y
Linear saddle (2,0.8); approximate q−invariant measure α=0.6
x
y
Linear saddle (2,0.8); approximate q−invariant measure α=0.75
x
y

Thankyou!

Abstract

In 1960 Ulam proposed discretising the Perron-Frobenius operator for a non-singular map (T,X) by projecting $L^1(X)$ onto the subspace of piecewise constant functions with respect to a fixed partition of subsets of X. Ulam's conjecture was that as the partition is refined, the fixed points of the approximation scheme should converge in L^1 to a fixed point of the Frobenius-Perron operator. Thus "Ulam's method" was born! Li (1976) proved the conjecture for piecewise C^2 expanding interval maps, and further results have been obtained by many authors over the subsequent decades. It is now clear that most of these results rely on strong analytical control of the spectrum of the Frobenius-Perron operator on suitable Banach spaces embedded in L^1. Indeed, in such settings, useful convergence rates can be obtained (for example, by using the spectral perturbation machinery of Keller and Liverani). However, applying these results to new classes of maps can be extremely difficult (or impossible); this is especially true examples coming from real applications. In this sense, a satisfactory proof of Ulam's conjecture remains elusive.

This talk will survey the ideas above, and describe a variational framework in which Ulam's method arises as one possible approximation scheme (joint work with C Bose). Analytical proofs of convergence can come "cheaply" in a variety of settings where the spectral perturbation approach does not apply, including some open systems. In keeping with the set-oriented theme of the conference, implementation relies on being able to compute intersections of elements of a partition of X, and topological features of the intersections turn out to be of great importance for the feasibility of the methods.

References

- Borwein & Lewis (1991) *Trans AMS* **325**:249-271
- Borwein & Lewis (1991) *SIAM J Control Optim* **26**:325-338
- Bose & Murray (2001) *DCDS* **7**:219-235
- Bose & Murray (2006) *DCDS* **14**:597-615
- Bose & Murray (2007) *SIAM J Optim* **18**:691-709
- Ding (1998) *App. Math. Comp.* **93**:155-168
- Ding & Zhou (1996) *Physica D* **92**:61-68
- Gora & Boyarski (1989) *Israel J Math* **67**:272-286
- Keller (1982) *Monatsh. Mathematik* **94**:313-333
- Keller & Liverani (1999) *Ann. Scuola. Norm. Pisa Cl. Sci* **28**:141-52
- Lasota & Mackey (1994) *Chaos, fractals and noise* (2nd ed), Springer
- Lasota & Yorke (1973) *Trans AMS* **186**:481–488
- Li (1976) *J. Approx. Th.* **17**:177-186
- Ulam (1960) *Problems in modern mathematics*, Interscience publishers.

A1: Explicit definition of quasi-compactness

- Let $\mathcal{B}$ be a Banach space, $\mathcal{P} : \mathcal{B} \to \mathcal{B}$.
  - spec(P) $\subset C$ is the *spectrum* of $\mathcal{P}$
  - $R = \sup\{|z| : z \in$ spec($\mathcal{P}$)$\}$ is the *spectral radius*
  - $\Sigma_0 = \{\lambda \in$ spec($\mathcal{P}$) $: |\lambda| = R\}$ is the *peripheral spectrum*
- $\mathcal{P}$ is *quasi-compact* if it decomposes as $\mathcal{P} = \mathcal{P}_0 + \mathcal{P}_1$ where
  - $0 \neq \mathcal{P}_0$ is finite rank, projecting onto the eigenfunctions for Σ_0
  - $\lim_{n\to\infty} \|(P_1)^n\|_{\mathcal{B}}{}^{1/n} = r < R$
- so Σ_0 is separated from the rest of spec($\mathcal{P}$): "$\mathcal{P}$ has a spectral gap"
- Usually $\mathcal{B}$ embeds in L^1, $\|\mathcal{P}\|_{L^1} = R = 1$, $\mathcal{P}$ **not** quasi-compact in L^1

  [there is an art to finding a $\mathcal{B}$ in which $\mathcal{P}$ is quasi-compact]